# *Moving Object's Detect in a Monocular Moving Camera*

CHEN Ye-gang
School of mathematics and computer
Yangtze normal university
Fuling,china

DIAO Xiao-rong
Fuling middle school
Fuling education Commission
Fuling,china

***Abstract*—Recently, there has been increasing interest in using mobile robots within spaces where humans reside, and safe navigation by effective sensing becomes an important issue. In this paper, we describe a method to detect moving objects that exist in front of an autonomously navigating robot by analyzing images of a monocular color camera mounted on the robot. In particular, detecting humans who walk in a direction opposite to the robot's motion is studied because this increases the danger of collision between the pedestrians and the robot. Although moving object detection and obstacle avoidance have been actively studied in the fields of computer vision and intelligent robotics, respectively, analyzing the images of a moving camera is still challenging. One method presented in this paper is based on comparing the current image and a past image in order to find moving objects in the scene. Assuming that the speed of the robot is known, a correspondence between a certain part of the current image and a matching part of a past image is established. Approaching objects can then be detected for a case in which the degree of mismatch between the two corresponding image parts is high. Another method we employ is the detection of human faces. Since a human face has unique features in color and shape, we can search faces in images in order to detect approaching humans. We propose a fast and simple masking method for face detection in a small search region specified from appearance-based foot detection. These two proposed methods were combined to effectively find approaching humans in our experiments. We could get promising test results.**

Keywords—***Autonomous smart robot; Moving camera; Human detection; Obstacle avoidance***

## I. INTRODUCTION

Computer-controlled reprogrammable machines are useful for flexible manufacturing, and robots are a typical reprogrammable machine. Industrial robots, which are often called manipulators, have been successfully used in production factories, particularly in automotive manufacturing, since the 1960s. Recently, thanks to rapid advances in computers, microelectronics, and mechatronics, robots have been employed outside of the factory, particularly for providing services to humans. Current service robots are often designed to be mobile and smart in uncontrolled environments such as home [1], office [2], and public places [3][4].

For autonomous navigation in an uncontrolled environment, a mobile robot should have a capability for dynamic path planning and obstacle avoidance with real-time computing [5][6][7]. If the workspace of a mobile robot coincides with that of humans (e.g., a service robot), the most prominent entities requiring the robot's attention during its motion are humans. In this situation, the safety of both the humans and the robot is the most important issue. The robot must be able to avoid any harmful situation to humans in addition to protecting itself.

Sensing obstacles is a fundamental function for a smart mobile robot in order to avoid collisions during its autonomous navigation. There are a number of sensors that can be used for obstacle detection. Laser scanners [8] provide accurate and dense range data around a robot. Time-of-flight (TOF) cameras [9] have many practical advantages because high resolution depth data can be obtained in real time with a single camera. However, laser scanners and TOF cameras are too expensive to be used to low-cost commercial robots. Ultrasonic range finders [10] are inexpensive and commonly used for the safe navigation of mobile robots, but the measurement resolution is very low. Infrared-based sensors [11] are cheap, and the measurement resolution is higher compared to ultrasonic range finders. However, the sensible range is short. Considering the different features of different sensors, fusing multiple sensors has been actively studied for various applications including robotic navigation [12].

Vision is the most versatile sensing capability in both men and machines. In recent years, the price of vision cameras has been falling rapidly, while their performance has been increasing. There are various low cost cameras that can be employed easily in many applications. Cameras that can be connected via a universal serial bus (USB) [13] are one example.

In this paper, we present practical methods to detect moving objects in the images of a camera fixed to a mobile robot. Although stereo vision has long been actively studied for the use with robots in order to perceive surroundings [14], it needs accurate mechanical setup and has high computational cost. Instead of using stereo cameras, we examined the visual detection of obstacles by analyzing the images of a monocular color camera mounted in front of a mobile robot. There are several practical approaches in detecting moving objects using a stationary camera, such as image differencing and background subtraction [15][16]. However, if a camera is attached to a moving robot, the problem of visual detection of moving objects becomes difficult because the ego-motion of the robot induces an inherent optical flow of the moving objects [17]. We propose two novel methods to detect moving

978-1-4799-3381-5/13 $31.00 © 2013 IEEE
DOI 10.1109/DASC.2013.96

objects in an indoor scene using a moving camera. Particularly, people walking in the direction opposite to the moving direction of a robot are detected because humans are usually the only moving entities, except the robot, in an indoor space. Assuming that the velocity of the robot is constant and known, some part of the current image and the corresponding part of a past image are compared to differentiate between static objects and moving people. Another method we examined is face detection. Since human faces have unique features in terms of color and shape, they can be detected in order to sense humans approaching the robot. We propose a simple mask-based face detector.

## II. DETECTING MOVING OBJECTS IN THE IMAGES OF A MOVING CAMERA

It is often required to automatically detect moving objects in the images of a camera. Examples include detecting people in a restricted area using a surveillance camera for a security reason, and monitoring vehicles on a road for traffic control. In the computer vision community, the background subtraction technique has been widely used for the detection of moving objects [15][16][17]. If a pixel value is monitored over time in a static scene, the value can be modeled with a Normal distribution, $N(\mu,\sigma 2)$, and any pixel outside the distribution model can be regarded as belonging to a moving object. A simple implementation of this method is to calculate an average image of the scene with no moving objects, subtract each new image frame from the average image, and threshold the result by comparing it with a value $k\sigma$, where k is a constant [17]. The background subtraction method is simple to program and fast to run. However, the method is not applicable to the images of a moving camera, which we consider in this study by attaching a color camera to a mobile robot.

When a camera is moving, the moving object detection problem becomes more complex than in the case of a stationary camera. However, employing a sophisticated algorithm cannot be a solution in practice. Vision for mobile robots has features that are similar to other data processing tasks in mobile devices where hardware and software resources are limited. A software algorithm running on a mobile system should be simple and light enough to run reliably with low-cost and low-performance devices. The appearance-based monocular color vision method of obstacle detection proposed in [18] is simple and can be easily implemented in a mobile robot. In this method, any pixel that differs in appearance from the ground is regarded as belonging to obstacles. The appearance-based method uses three assumptions: (i) obstacles differ in appearance from the ground, (ii) the ground is relatively flat, and (iii) there are no overhanging obstacles. These assumptions are reasonable for a variety of environments, and we used the appearance-based method to detect obstacles. First, a camera image in red-green-blue (R-G-B) color space is transformed to the corresponding hue-saturation-value (H-S-V) space to separate color and intensity. Then, pixels having different color values from those of the ground are detected. To simplify the detection process, the detection is made only in a certain region that is set horizontally in the image. If there is an approaching person in front of the camera, a body parts that first appear in the horizontal search region are one's feet as the distance between the person and robot becomes smaller. To reduce the influence of surrounding light, the detection is made by comparing only H and S color values.

The aforementioned appearance-based method detects pixels having different color values from those of the ground. Thus, the detected pixels can belong to any objects that are static or dynamic. However, the main goal of this study is the detection of moving objects. For the case of stationary objects, their locations are known beforehand and they already exist in the map of a navigating robot. Thus, in practice, higher importance should be given to the detection of moving objects. We propose a method based on the comparison of two images: a past image and the current image. When a camera is mounted on a robot that navigates at a constant speed, it is possible to estimate where the objects in the current image will appear in the next image. In the same way, we can estimate where current objects were in a past image. Figure 1 shows an example. When a robot moves on a corridor floor, a point pT at a row coordinate rT in the current image corresponds to pT-N at a row coordinate rT-N in a past image that was captured at time t=T-N. This is an ideal situation. In practice, as shown in Figure 2, the motion of an image point due to the camera's motion will not only be vertical. Furthermore, the distance $\Delta p$ between pT and pT-N due to the camera's motion varies depending on the position of the point in the image because of the optical and geometric characteristics of the camera. However, the basic idea of the ideal situation is still useful even in practice. The vertical travel distance of a point $\Delta pr$ is much larger than the horizontal distance $\Delta pc$. If the frame rate of a camera is high, $\Delta pr$ can be small. Then, $\Delta pc$ is even smaller if the robot does not change direction during navigation. Therefore, we can check the correspondence between the current and past images only vertically. If there is an object moving in the direction opposite to the robot's motion, a serious mismatch will occur between the two compared image lines.

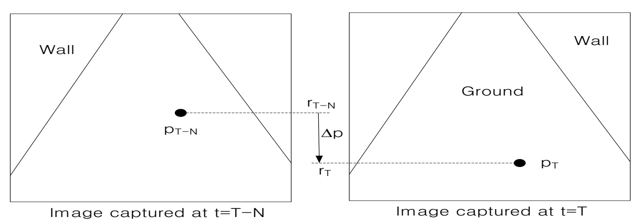


Figure 1. Ideal displacement of a point due to camera motion.

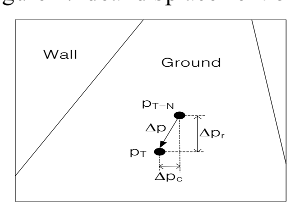


Figure 2. Practical point displacement.

The method that we propose is summarized in Figure 3. For a current image I(T) and a past image I(T-N), the color images in R-G-B space are transformed to those in H-S-V space, resulting in images J(T) and J(T-N). A belt-shape image called L1 along a detection line is sampled as J(T), and the corresponding image L2 is sampled as J(T-N), as shown in Figure 4. The vertical difference between L1 and L2, $\Delta L$, is set in advance by considering the speed of the camera motion (i.e., the robot's speed). The width of the ground line in each image is determined by detecting the point at which the ground meets the wall. The width of L1 is w1, and h pixels above and below the scan line are taken into account in the comparison. Note that the size of image L1 is then $(2h+1)\times w1$ pixels and the image size of L2 is $(2h+1)\times w2$. These images are resized to smaller line images U1 and U2, which are in $1\times w$ pixels, using the nearest-neighbor interpolation method. The image size reduction can decrease the effects of image noise, inaccurate setup, and error in the constant robot speed. Then, the intensity values (i.e., V values in the H-S-V color space) of the two line images U1 and U2 are compared. Since the two images correspond to the same ground line in the scene, they should, ideally, be the same. However, if there is a moving object such as an approaching person, the feet appear in U1, and this will cause significant values in the difference image D.

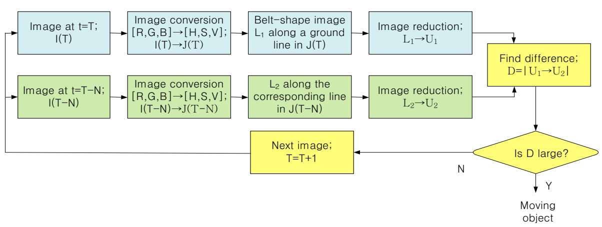


Figure 3. Moving object detection by comparing two images.

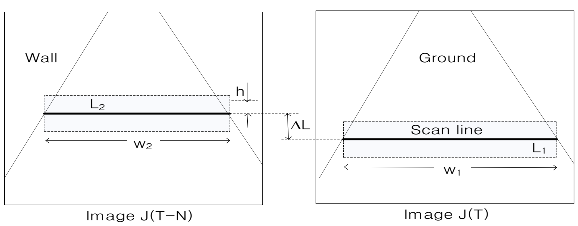


Figure 4. Comparison of current and past images along a scan line.

## III. FINDING FACES TO DETECT APPROACHING PEOPLE

Humans usually expose their faces, unlike most other body parts that are covered by clothes of various shapes and colors. Therefore, faces are comparatively easy to detect in images of humans. Additionally, the human face has unique features that can be predictable in terms of color, shape, and position (relative to other body parts). We detect faces in images to determine whether people are approaching the robot.

Among the current face detection techniques available in the computer vision community, the most popular method might be that proposed by Viola and Jones [19], which uses the AdaBoost (short for Adaptive Boosting) algorithm. However, the AdaBoost method requires rather a large image resolution (e.g., 24×24 pixels). Since this can bring a practical difficulty to low-cost commercial robots, processing with low-resolution input images is strongly recommended [20]. Processing time is another practical factor to consider. Sophisticated face detectors are often computer intensive, which is undesirable for a mobile system.

In this study, we use the color of the facial skin for the detection. The skin color of a human forms a dense distribution as exampled in Figure 5. This unique skin color distribution is an important clue in the detection of human faces in an image. If a camera is stationary, the face can be the largest skin-colored moving object in the images [21]. However, in this study, the camera was assumed to be moving: thus, detection based only on skin color will not bring accurate results. We propose a simple face detection method using a mask, as shown in Figure 6. The mask consists of three parts: (B), (H), and (F). Part (F) is the largest of the three parts representing the face skin. Part (B) is the smallest part, and represents the background behind the head. Part (H) represents hair, and is located between the other two parts. The pixels of Part (H) can be assumed to be very dark for most Asian people, whereas the pixels of Part (B) should not be dark. The H and S color values of pixels in Part (F) must follow the distribution of skin color, which we can define beforehand from sample face images. In a masked image, we count pixels that belong to the skin color distribution within Part (F), and the count will be maximized if the part is best overlapped with a face. The size of Part (F) is determined by considering image resolution. The pixels of Part (H) are counted if their V value is quite low (nearly black), whereas the pixels of Part (B) are counted if their V value is moderate or high (considering the brightness of normal indoor environments). The size of each part indicates the weight in searching - the largest size of Part (F) means that we use the skin color as the most important clue in the search. The total counted number of pixels satisfying the intensity and color conditions in the three parts is used as a matching score. This simple method of using the mask effectively takes into account the color, intensity, and topology of a human face. The face detection mask is applied to a small search region that is defined for a certain distance above the feet. Note that the feet were detected in the image by the appearance-based method described in Section 2. At each position within the search region, the score is counted, and the position of the highest score is estimated to belong to a face. The center of the mask is set at the center of Part (F). However, if the highest score counted is lower than a threshold, we decide that there is no face. Figure 7 summarizes the proposed face detection method.

ou begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

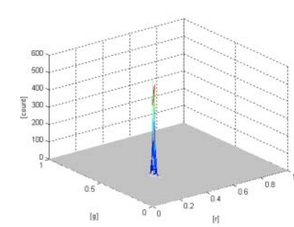

Figure 5. Dense color distribution of human skin

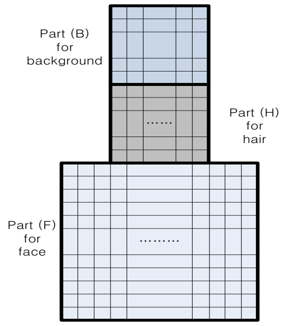


**Figure 6**. Mask for detecting face.

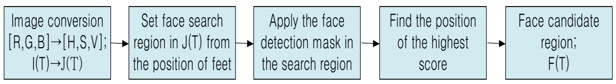


**Figure 7**. Face detection in an image.

## IV. RESULTSTE

The proposed methods were tested using a color camera mounted on a wheeled mobile robot that navigated with a constant velocity in a building corridor. Figure 8 shows examples of our moving object detection technique. Figure 8(a) shows some experimental scenes. All of the test images have a resolution of 480×640 pixels. In Scenes 1 and 2, there is an approaching human. In Scene 3, a wooden box (and nothing else) is located on the right side of the floor. The main goal is to distinguish between dynamic obstacles and static objects, particularly in order to detect moving humans who walk in the direction opposite to the moving robot.

First, obstacles on the corridor floor, regardless of whether they are static or dynamic, are detected by the appearance-based method described in Section 2. Figure 8(b) shows the detection results for the experimental scene images. The top image shows the search region of an image, the middle image shows the pixels having different colors from those of the floor, and the bottom image shows the final detection result. The final result was obtained by projecting the detected pixels onto the horizontal axis, and removing detected areas that were too small or too large. When a human approaches the robot, the body parts that first appear in the band-like search image region are the feet. In Scenes 1 and 2, the feet of the pedestrian were readily detected.

Secondly, the face is searched in order to decide if the obstacle detected by the appearance-based method is a human. The face detection results are presented in Figure 8(c). The detection mask shown in Figure 6 was applied to a search region that was set above the part detected by the appearance-based detection process shown in Figure 8(b). The left image shows the result of masking, and the right image shows the detected face box at the highest masking score in the search region. The pedestrian's face was readily detected in both Scene 1 and Scene 2. In the image of Scene 3, no face was detected although an obstacle was detected in Figure 8(b). This could mean that the detected obstacle was not a human.

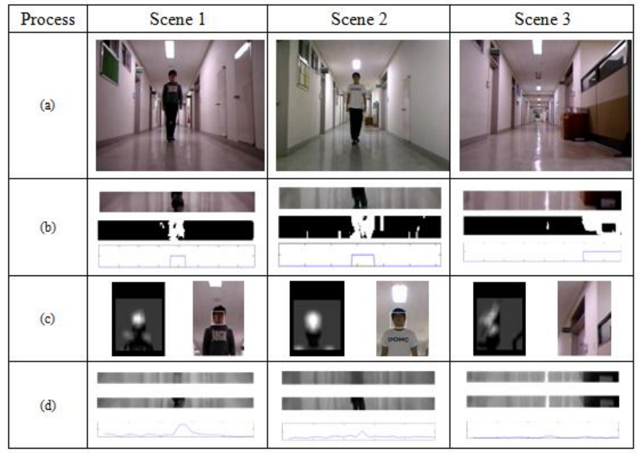


Figure 8. Experimental result. (a) Original scene images, (b) Appearance-based obstacle detection, (c) Human face detection, (d) Moving object detection.

Walking humans (in Scenes 1 and 2) can be distinguished from the static wooden box (in Scene 3) by motion detection, which produced the results shown in Figure 8(d). A moving object was detected by comparing the corresponding scan lines in past and current images, as described in Section 2. An eleven-pixel-high belt-shaped image that centers the scan line was resized into 1×25 pixels for a simple and robust comparison. In Figure 8(d), the top and middle images represent two compared images of the same position on the floor. The bottom image is the average intensity difference of the two compared images. In the case of an approaching human, a large difference occurred at the human's position as shown in Scenes 1 and 2. In comparison, the static wooden box in Scene 3 did not result in a significant difference.

## V. CONCLUSION

As robots become mobile and smart, their safe navigation becomes an important issue. We propose techniques for detecting moving objects, particularly people approaching a mobile robot, using a low-cost monocular color camera mounted on the robot. Since everything appears to move in the sequential images of the moving camera, existing methods for a stationary camera, such as background subtraction and image differencing, cannot be used in this situation. We compared a current image and a past image with respect to corresponding scan lines in the images to find moving objects. Our experimental results showed that the proposed method could distinguish between dynamic and static objects in a scene. Faces are detected to find approaching humans because a human is the most important entity for a mobile service robot to recognize. Many of existing face detectors are computer intensive or requiring high-resolution images that are difficult to apply to low-cost mobile systems including commercial robots. Thus, we devised a simple mask-based method. By applying a mask on a search region, we could find a face quickly at the highest masking score of the image. Since the proposed methods are simple, they can be readily used with many commercial mobile robots..

## REFERENCES


[1] F. Lu, G. Tian, X. Wang, "Intelligent space system oriented to the home service robot", in Proc. Int. Conf. System Science and Engineering, pp. 50-54, 2012.

[2] C.-F. Wei, Z.-W. Lin, K.-Y. Lian K-Y, "Development of an office delivery robot with multimodal human-robot interactions", .in Proc. SICE Annual Conf., pp. 1564-1567, 2012.

[3] F. Capezio et al., "The ANSER project: Airport nonstop surveillance expert robot", in Proc. IROS, pp. 991-996, 2007.

[4] J.-G. Kang, S.-Y. An, S.-Y. Oh, "Navigation strategy for the service robot in the elevator environment", in Proc. Int. Conf. Control, Automation and Systems, pp. 1092–1097, 2007.

[5] Z. Wu, L. Feng, "Obstacle prediction-based dynamic path planning for a mobile robot," Int. J. Advancements in Computing Technology, vol. 4, no. 3, pp. 118-124, 2012.

[6] B. I. Kazem, A. H. Hamad, M. M. Mozael, "Modified vector field histogram with a neural network learning model for mobile robot path planning and obstacle avoidance", Int. J. Advancements in Computing Technology, vol. 2, no. 5, pp. 166-173, 2010.

[7] M. K. Habib, "Real time mapping and dynamic navigation for mobile robots", Int. J. Advanced Robotic Systems, vol. 4, no. 3 , pp. 323-338, 2007.

[8] C. Ye, J. Borenstein, "Characterization of a 2-D laser scanner for mobile robot obstacle negotiation", in Proc. IEEE Int. Conf. Robotics and Automation, pp. 2512-2518, 2002.

[9] S. Hussmann, T. Liepert, "Robot vision system based on a 3D-TOF camera", in Proc. IEEE Instrumentation and Measurement Technology Conf. (IMTC 2007), pp. 1-5, 2007.

[10] H. R. Everett, Sensors for Mobile Robots: Theory and Application. Wellesley, Mass., A K Peters, 1995.

[11] Y. Do, J. Kim, "Infrared range sensor array for 3D sensing in robotic applications", Int. J. Advanced Robotic Systems, vol. 10, 2013.

[12] T. Chen et al., "Robot navigation strategy on kernel-methods-based multi-sensor fusion", Int. J. Advancements in Computing Technology, vol. 5, no. 3, pp. 756-769, 2013.

[13] C. Cardeira, J. S. da Costa, "A low cost mobile robot for engineering education", in Proc. 31st Annual Conf. IEEE Industrial Electronics Society (IECON 2005), pp. 2162-2167, 2005.

[14] X. Lai, H. Wang, Y. Xu, "A Real-time range finding system with binocular stereo vision", Int. J. Advanced Robotic Systems, vol. 9, pp. 1-9, 2012.

[15] Y. You, S. Gong, C. Liu, "Adaptive moving object detection algorithm based on background subtraction and motion estimation", Int. J. Advancements in Computing Technology, vol. 5, no. 6, pp. 357-363, 2013.

[16] V. Reddy, C. Sanderson, B. Lovell, "Robust foreground object segmentation via adaptive region-based background modeling", in Pro. Int. Conf. Pattern Recognition (ICPR), pp.3939-3942, 2010.

[17] A. Elgammal, D. Harwood, L. Davis, "Non-parametric model for background subtraction", in Proc. European Conf. Computer Vision (ECCV), pp. 751-767, 2000.

[18] I. Ulrich and I. Nourbakhsh, "Appearance-based obstacle detection with monocular color vision", Proc. AAAI, 2000.

[19] P. Viola, M. Jones, "Robust real-time face detection," Int. J. Computer Vision, vol.57, no.2, pp.137-154, 2004.

[20] D. Kim et al., "Vision-based arm gesture recognition for a long-range human–robot interaction", pp. 1-17, J. of Supercomputing, 2011.

[21] B. Schiele, A. Waibel, "Gaze tracking based on face-color", in Proc. Int. Workshop Automatic Face- and Gesture-Recognition, 1995.